STATE UNIVERSITY OF NEW YORK AT ALBANY
1400 Washington Avenue, Albany, N. Y. 12203

FUNCTIONAL ANALYSIS OF INFORMATION RETRIEVAL

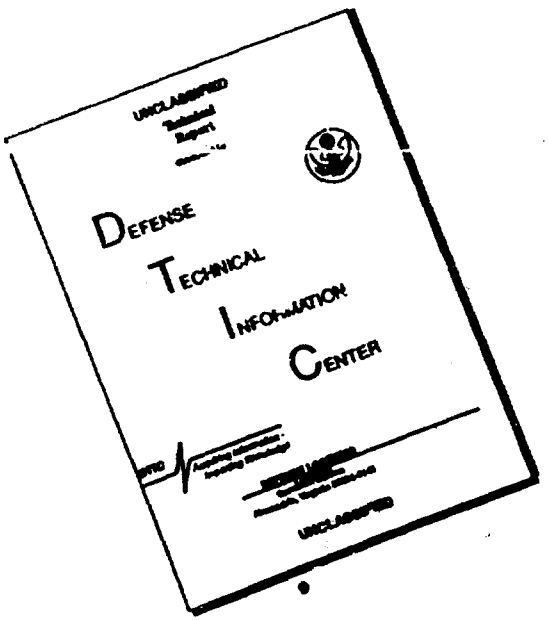
by
Robert A. Fairthorne

FINAL REPORT

August 1968

This report was prepared under the sponsorship of
the Air Force Office of Scientific Research,
Directorate of Information Sciences, Grant AF-AFOSR-68-1421

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

FUNCTIONAL ANALYSIS OF INFORMATION RETRIEVAL

The purpose of this final report is to summarize the content of a series of lectures developed for the curriculum of the School of Library Science, State University of New York at Albany.

1. The Limits of Information Retrieval.

This lecture, prepared for a general audience, discusses the need for a (correct) theory of documentation, and how it can be derived. First, instead of trying for a comprehensive theory of all documentation and of its applications, one should examine an activity that is necessary for all such activities, even if not sufficient. Information Retrieval, which is the activity concerned with supplying readers with the kinds of messages they prescribe, is necessary to all documentary services. If one cannot do this, one cannot do the others.

Thus we confine ourselves to the necessary. Also we must exclude the impossible. In every field there are Principles of Impotence; things that cannot be done, except by chance, however ingenious and hard working the methods and agents. These principles are powerful tools (cf. thermodynamics).

Clearly no aspects of Information Retrieval should entail omnipotence, nor should they be tested against imagined omnipotence. In particular, Information Retrieval has no control over who writes what, or how well; nor upon who asks for what and how he uses it when he gets it.

Like all disciplines dealing with symbolic representations, it is limited by the Principles of Ignorance: "You can't tell whether a statement is true by looking at it or at computations performed upon it."

Hence, for instance, the folly of assessing retrieval performance by how 'useful' the output is to the reader. Not only does this depend upon authors and what they write, but it would require a system that could separate true statements from false. If we had this, we could dispense with experiment. Instead we would write various contradictory assertions about the matter in question, and the system would select the true one.

Thus retrieval services in particular, and library services in general, must be judged by how well they serve the readers in getting what the reader prescribes. They must not be judged by how well or desirably the reader makes use of what he gets, within library regulations.

The only facts that an information retrieval service can retrieve are facts concerning discourse, not concerning what the discourse is about. That is: it retrieves factual statements at best. To verify such statements as are verifiable - many are not, because they refer to fictions - one must go outside the field of documentation, with suitable instrumentation.

Before one can retrieve a recorded statement, one must retrieve the document containing it. A document is a unit of recorded discourse natural to the local environment of discourse. That is, a document is a document if those making use of it say it is a document.

Discourse is important, so recorded discourse is important. If it were not, there would be no need for libraries. But one must not confuse what things are with people have to say about them.

Information Retrieval must deal with both physical and linguistic matters. Documents are heavy objects with marks on. Their interpretations are not. A documentary situation may be purely physical; i.e. it is what it is, however people talk about it, if at all. At the other extreme it may be purely social; i.e. it is what the people concerned say it is, or it isn't anything at all. In most documentary situations the physical and social components need careful sorting out.

Too much confusion about what is otherwise obvious in this field arises from confusing what is spoken about with how it is spoken about. Consider these three sentences:

"A gives B information."
"A informs B about C."
"A tells B about C."

The first implies transfer of some self-subsistent substance called "information." As a metaphor, it may pass. As an image of reality it has proved disastrous.

The second implied that B's knowledge about C is changed by what A has said or written. Whether this is so depends on many factors outside the control of Libraries and librarians as such; for instance what A has said, and B's personal history and psychology. Even if B's knowledge has changed, it may have changed for better or worse.

The third is the only level at which librarians can work as librarians. They can make it easier for B to find out about A, and others, have to tell about C.

The aim is big enough to keep librarians busy indefinitely without taking over other people's professions and jobs. It is also an aim worthy enough to be called a profession in its own right.

2. The Structure of Notification. Informal talk and discussion in the School of Library Science, SUNYA, 13 December 1967.

This talk and discussion considered in more detail matters raised in the 'Limits' lecture. The main weapon was the analysis of operations introduced in the paper 'Morphology of "Information Flow"', published in JACM, October 1967 (see Appendix). This was introduced in ab initio in terms more appropriate to a library school.

First, 'Notification' was shown to be the smallest field that exhibits the essential features of library work, without being reduced so much as to fall below this field i.e. it did not exhibit the fallacy of regarding one man as being a small crowd. 'Notification' was defined as the service that relates a Destination to the Messages (if any) that he or it prescribes. Thus it is the delegate of the Destination, not its substitute. Notification cannot possibly reproduce the behavior of an individual selecting documents from the shelves unless that individual can state how he selects certain documents and not others. That is, the Destination must state, or be brought to state, the sort of messages he requires in terms intelligible and usable by someone else.

According to the analysis of the paper cited above, his prescription must be in terms of Source, Code, Channel, and Designation. Here Designation can be taken as the 'aboutness' of the message relative to the environment of use. These elements can realize themselves in many ways, and each has (as pointe out by Calvin Mooers) qualitative and quantitative aspects. The discussion clarified these points, and then centered on identification of the various triads of elements involved in Notificational activities; i.e. the sixteen triads that do not contain both Destination and Message.

Many apparently distinct activities were found to employ the same three elements, and general enough names for these were difficult to devise. For any three elements variations occurred according to which of these were regarded as given, and which variable e.g. within the 'Shannon' triad of Message, Code, Channel, fixing of channel characteristics and message use statistics gives 'detection of signals in the presence of noise'.

Synthesis of these triads to form more familiar but composite operations gives six fundamental patterns (tetrads) corresponding to distinct families of notificational methods.

The discussion of the six elements and their combinations, particularly with respect to identification with library and documentary activities and interests, proved this line of attack to be powerful and educational.

(It was clear also that even those with some mathematical background find some difficulty in thinking in terms of functions of more than one variable, as triads and tetrads must be. Much grief in the devising of 'measures' for aspects of retrieval performance comes from pathetic attempts to compress a funtion of at least two variables into a function of one. The example of thermodynamics should be a clear enough warning.)

3. About 'Aboutness'. Lecture and discussion in the School of Library Science. SUNYA, 30 April 1968.

Aboutness has been discussed from Plato through Nelson Goodman. Here only the documentary aspects were discussed; that is, the aboutness of records inasmuch as it affects their retrieval. This is fundamental to all library services. Librarians are expected to know about discourse, and what to do about it, but not to take part in it.

'Aboutness' is not a simple, or even a unique characteristic of documents. Clearly parts of a document, in isolation, may not be about what the entire document is about (e.g. conversations in fiction, parts of a reference book.). Nor is a document about the sum of the things that its statements are about Aboutness is not a thing or an intrinsic property, but a relation. Therefore there are many kinds of aboutness, depending upon the social environment of a document's use or production. As in all social situations, one must ask not only 'how?' and 'what?', but also 'why?'; a question that does not arise in the natural sciences.

Documentary aboutness may be the Designation of a Message with respect to Code, Source or Destination. Library services exist only to serve the last, the reader. In general, Destination and Source will require different Designations. They will be the same only when Destination and Source are members of the same group of peers, e.g. scientists or technicians, writing for each others benefit (L. M. Bohnert); when the topics are 'general' in the sense of being commonly understood in the social environment concerned or when the document is 'popular', the Source identifying himself with the Desintation.

The triad Message, Code, Designation arises in authorship rather than in librarianship. Indexing procedures based on uninterpreted text ('Mechnaical Indexing') do not deal with the Message, so may deal with either triads, Code, Designation, Source, or Code, Designation, Destination.

Whatever the triad in which it occurs, aboutness (Designation) can be extensional or intensional. From the extnesional point of view of view a document is about what it mentions, e.g. 'Moby Dick' is about a whale, amongst other things. Problems here are to determine what is mentioned, the role of a document as a complex assertion, and attributive assertions. Because a document is a unit of discourse, not a mere assemblage of linguistic expressions, coexistence of assertions within a document has special significance (D. J. Hillman).

This significance is the intensional aboutness. For instance, the Encyclopedia Britannica is about (extensionally) the matters it refers to in its various articles. Intensionally it is about most of the things that most educated English speaking people want to know about some time or other. This is why it was compiled, published and purchased. Intensional aboutness is intimately connected with acquisition policy. It is established by 'matching' (L.M. Bohnert) against documents that are 'similar' with respect to the sort of requests that are expected, or against dissemination policy, e.g. who reviews or abstracts a paper is settled by the editor of the review journal.

Intensional aboutness cannot be determined from the text, or its interpretation, alone, because one cannot deduce the question from its answer. If one could, one could dispense with experiment, and violate the Principle of Documentary Ignorance. To determine it, one must consider how the document might be modified by contradictions and substituions, and still be considered to be about the same thing in the given environment. For 'aboutness', extensional or intensional, entails ignorance in the sense that it discriminates at a corser level than the intending reader requires. Otherwise the intending reader couls not ask for the document before seeing it.

Consider, as an example, a document whose main argument was that 'X is a blithering idiot'. If a document whose argument was 'X is not a blithering idiot' would have the same aboutness, than both documents are about the blithering idiocy of X. But if documents of the sam aboutness had arguments of the kind 'Y is a blithering idiot', 'Z is a blithering idiot', 'Miss A. is a blithering idiot', these documents are about the distribution of blithering idiocy amongst a subsection of the human race.

To sum up, 'aboutness' (Designation) of documents can be intensional or extensional, and differs according to its relation to circumstances of production, and circumstances of use. In general it cannot be determined by consideration of textual characteristics alone. It is not concerned with the truth, falsehood, or logical consistency of the statements made in the document.

4. 'Information Sciences' and the Library School Curriculum. Discussion at Faculty Meeting, School of Library Science, SUNYA, 27 May 1968.

The key topics of the discussion were the nature of the 'Information Sciences' and the nature of Library Science and Library Service. Without some agreement on these, their interactions could not be discussed usefully. My view of the "Information Sciences' is that stated in my FID paper "The Scope and Aims of the Information Sciences and Technologies' (see Appendix). In brief, this holds that the term applies to various applications of diverse technologies and skills to facilitate discourse. Their only common feature is in that application. There is no more need for those who facilitate discourse to understand these diverse sciences and technologies, then there is for those who asks omelettes to be able to lay eggs. Or, to take a more cognate example, for those who use telephones for communication to understand electomagnetism or acoustics.

What is required is knowledge and understanding of what these technologies can do, at a given time and for given cost, within the field of application. For example, librarians have something to learn from computing people about the management of large files of certain kinds. They can do so without having the slightest knowledge of computers as such; just as they can make use of computers (or railroads, or postal services, or telephones) without having any knowledge of computers (or railroads, or postal services, or telephones) as such, so long as they know what these entities can do in terms of their intended application.

Similarly those who do understand how computers, say, work must think in terms of the intended application. When computers are technically capable of being used to store and manipulate customers' records, they must be regarded as libraries and thought of in library terms. In this particular example the essentials are to allow the used to work the system by himself; to preserve records from selfish, malicious or careless actions; and to allow for growth and change of demand. These are not principles of computer science, but of library (or managerial) sciences.

The library sciences are essentially those associated with helping people find out what people have had to say; more shortly, with aiding discourse, particularly recorded discourse, on behalf of the recipient, not the originator, of the discourse. Thus it is inverse to reprography, signalling, and computing, which work on behalf of the originator. But librarianship should make use of all these, on its own terms.

The library school curriculum could make use of these 'Information' technologies if, first, they were presented in terms appropriate to library activities and ideas; second, the terminology itself had some relation to well defined ideas, was reasonably unambiguous and was commonly agreed upon some degree of literacy would help.

Some of the discussants expressesnt discomfort with all these points, though none disagreed with all of them. The main difficulty was the idea of a librarian being a specialist in discourse, but not a participant in it: i.e.if he took an active part in discourse, he ceased to be acting as a librarian. This was not due to a belief that librarians must be omniscient in the sense of knowing about all the topics their documents were about. It arose more from the great difficulty of distinguishing between 'book knowledge', and understanding and exerience of a topic. Not all that is said is knowledge, nor can all knowledge be said, e.g. Noone can write a useful text on 'How to learn one's first language'.

Some thought that to define a librarians job as 'to aid discourse on behalf of the reader (not as a substitute for the reader)' too narrow. However, aiding discourse involves more than directing appropriate and available documents to those who request them. To do it adequately a librarian must be expert in some special kind of discourse, in the sense of knowing who writes about what, what has been written about what, what terms it is written in and asked for, what sorts of people ask for what sorts of documents, and how to get hold of these documents and people. Also he is not forbidden to infludence terminology, format, style and other bibliographical and literary matters that hinder discourse. But he is not expected to be, nor can he be also a full time practioner and expert in the topics this discourse is about. A librarian's job is to do things for the reader that the reader cannot reasonably be expected to do or to arrange for himself. It is not to teach the reader the reader's job.

Conversely it is not for the computer specialist, or mathematician, or whatever, to teach the librarian the librarian's job, but to help him do it.

5. Other Activities. The lectures summarized above arose from and led to other discussions and activities. Formal presentations are listed in the Appendix.

Visitis were made to the National Bureau of Standards, University of Maryland, and Case Western Reserve University, Cleveland, Ohio. I attended the Annual Meeting of American Documentation Institute (now ASIS) in New York City, October 1967. There I was awardded the Annual Award of Merit for 1967.

By invitation I made an informal presentation, about 'Aboutness', to a private discussion group of assorted scientists (astronomy to chemistry) and historians of the Rensselaer Polytechnic Institute, Troy, New York. This meets periodically to discuss aspects of the philosophy of science.

I continue to review papers of 'informational' imports for 'Computing Reviews'.

6. Acknowledgements. The Research Foundation of the State University of New York and the Office of Sponsored Funds, SUNYA, have gone out of their ways to deal smoothly with both routine and special financial matters.

Dean Farley, the Faculty and Staff of the School of Library Science, SUNYA, not only supported the rather distracting presence of a Visiting Research Professor with a great deal of work outside the normal call of duty, but also made me very welcome.

Though I retrain the right to claim intellectual errors in my work as my own contribution, much of what is not in error should be credited to the Faculty and Students of the School of Library Science. They have done their best to keep me up to the mark, and in a most friendly and pleasant manner.

In particular I must thank Mrs. L. M. Bohnert, Assistant Professor, and Mr. Jackson David, my Graduate Assistant. Not only did they compensate for my lack of competence in clerical and administrative matters. More important, they refused to take any assertion of mine for granted, and posed questions to me that did not allow slurring over essential difficulties.

Very sincerely yours,

Robert A. Fairthorne

Research Grant AF 68 1421

PUBLICATIONS AND PRESENTATIONS
October 1967 through June 1968

1. Publications

Morphology of 'Information Flow'.	J. Association for Computing Machinery, 14, 4, October 1967 pp. 710-719.
H. P. Luhn; applied mathematician.	in: Schultz, C.K. (ed) Hans Peter Luhn - Pioneer and Prophet of Information Processing. pp. 21-23 (Spartan Books. 1968)
Information Processing: History.	Encyclopedia Britannica. 200th Anniversary Edition 1968.
Critique of SOERGEL, D. 'Remarks on Information Languages.'	in International symposium on Relational Factors in Classification, June, 1966, University of Maryland. Information Storage and Retrieval. 3, 4, December 1967, pp. 293-294.
Essay-review of FARRADANE, et. al. 'Report on research in information retrieval by relational indexing...".	J. Documentation, 24, 2, June 1968, pp. 127-131.
The Scope and Aims of the Information Sciences and Technologies.	Invited paper for Committee FID/RI, 34th Meeting of F.I.D. Moscow, Sept. 1968. (to be published)
The Limits of Information Retrieval.	J. of Library History, Philosophy and Comparative Librarianship. (to be published, October 1968.)

2. Presentations.

The Limits of Information Retrieval.	Colloquim, State University of New York at Albany (SUNYA), 13 November 1967.
Structure of Notification	School of Library Science, SUNYA, 13 December 1967.
On "Question - answering' by Library Services	National Bureau of Standards, Technical Information Division, 11 January 1968.
Scope and Limits of Information Retrieval.	American Institutes for Research, Silver Spring, Md. 12 January 1968.
Scope and Limits of Library Services	School of Library Science, University of Maryland. 13 January 1968.
Technological Factors in Social Change	Radio and T. V. Association of N.Y. 15 April 1968.
About "Aboutness'.	School of Library Science, SUNYA, 30 April 1968.
'Information Sciences' and the Library School Curriculum.	Faculty Meeting, School of Library Science, SUNYA, 27 May 1968.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
State University of New York at Albany
School of Library Science
1400 Washington Avenue, Albany, New York 12203

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP

3. REPORT TITLE

FUNCTIONAL ANALYSIS OF INFORMATION RETRIEVAL

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Scientific Final

5. AUTHOR(S) (First name, middle initial, last name)

Robert A. Fairthorne

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
August 1968	9	0

8a. CONTRACT OR GRANT NO. AF-AFOSR-68-1421	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO 9769-02	
c. 6144501F	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d. 681304	

10. DISTRIBUTION STATEMENT

1. This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
TECH OTHER	Air Force Office of Scientific Research(SRI) 1400 Wilson Boulevard Arlington, Virginia 22209

13. ABSTRACT

The report summarizes the work under four headings. (1) The limits of Information Retrieval, which are set by the Principles of Documentary Impotence and Ignorance. The first asserts, inter alia, that librarians cannot supply documents that have not been written nor can dictate how documents will be used. The second asserts, inter alia, that one cannot tell whether a statement is true just by looking at it or at computations performed upon it. These principles have fruitful consequences. (2) The structure of retrieval services are analyzed and synthesized by identifying various instances of Source, Destination, Designation, Code, Channel, and Message, and the sixteen triads of these that do not include both Destination and Message. Identification varies according to which entities are given, which dependent. (3) Documentary 'aboutness' varies as Designation is related to Message with respect to Source, Destination, or Code. Extensional aboutness (what a document refers to) differs from Intensional Aboutness (why a document is about what it is about). (4) 'Information Sciences or Technologies' are largely applications of distinct techniques to documentary problems. If there be any common principles, these arise from their application to aid discourse, not from the disciplines from which these techniques originate. The report lists publications and presentations carried out under the Grant.

DD FORM 1 NOV 68 1473

UNCLASSIFIED

Security Classification

Key Words

Aboutness
Content analysis
Designation
Documentation
Education
Information sciences
Information retrieval
Librarianship
Notification
Relevance